

## SEQUENCE LISTING

SEQ ID NO:1

Human IC-RFX cDNA sequence

5           1   TTTCTGCGCT GAGCCAGGGC ACCCCGGAGC CTGCGGCCTC CTTCCCCGCC  
          51   CCTGCGGCC CCGGTCCCAG CCCC GCCCG CCCC GCCCCG GGCTGGGGCT  
10          101   CCGCTGGGA ACCGGCCGAG CGGCGCGC GGAGGTGTCC GGCGGCCAGG  
          151   AGGATGGCCA AGGTCCCAGA GCTGGAAGAC ACCTTCCTGC AGGCGCAGCC  
          201   TGCGCCCCAA CTGTCCCCGG GGATCCAGGA AGACTGCTGT GTGCAGCTCC  
15          251   TGGGCAAGGG CTTGCTAGTC TATCCGGAAG AACACAGTGT A CCTGGCGGCC  
          301   GAAGGGCAGC CCGGGGGCGA GCAGGGCGC GGGGAGAAAG GCGAAGACCC  
          351   GGAGCTGCCG GGGGCAGTGA AATCAGAAAT GCACTTAAAC AATGGTAAC  
20          401   TTTCCTCTGA AGAACAGGAC GCCGACAACC ACGACAGCAA AACCAAAGCA  
          451   GCGGATCAAT ACCTGTCTCA GAAGAAAACC ATCACGCAGA TTGTGAAGGA  
25          501   TAAAAAGAAG CAGACACAGC TCACGCTGCA GTGGCTTGAA GAGAATTACA  
          551   TTGTATGTGA AGGAGTTTGC TTACCACGGT GCATTCTTA TGACACACTAC  
          601   TTAGATTTCT GTAGGAAAGA GAAATTAGAG CCAGCCTGTG CGGCCACCTT  
          651   TGGAAAGACA ATTGCCAGA AGTTTCCCT CCTAACAAACA AGGCGGCTTG  
          701   GAACAAGAGG CCATTCAAAG TATCATTACT ATGGGATTGG CATCAAAGAG  
30          751   AGCAGTGCAT ATTACCACTC CGTTTATTCT GGAAAGGGCT TGACAAGGTT  
          801   TTCTGGAAGC AAGCTAAAGA ATGAGGGTGG CTTCACTCGT AAATATTCGC  
          851   TTAGCTAAA AACTGGAACA CTTCTTCCAG AATTCCCCAG CGCTAACAC  
40          901   CTTGTATACC AAGGATGCAT TTCTAAGGAC AAGGTTGATA CGCTCATAAT  
          951   GATGTACAAA ACTCACTGCC AGTGTATCCT GGACAATGCA ATTAATGGAA  
45        1001   ACTTTGAAGA GATCCAGCAT TTTTTATTAC ACTTTGGCA AGGAATGCCT  
          1051   GACCATCTCC TTCCCTGCT CGAAAATCCT GTTATCATTG ATATTTCTG  
          1101   TGTTTGTGAC TCAATTCTTT ATAAGGTTCT TACAGATGTA CTCATTCTG  
          1151   CAACAATGCA AGAAATGCCT GAAAGCTTAT TAGCAGACAT AAGAAATTT  
50        1201   GCTAAAAATT GGGAACAGTG GGTTGTTCA TCCTTGGAAA ACTTGCCAGA

1251 AGCTCTAACT GACAAGAAAA TACCTATTGT GCGAAGATT GTATCTTCTC  
1301 TGAAACGACA AACATCTTC TTACATCTTGC CCCAGATTGC CAGACCAGCT  
5 1351 CTCTTGACC AGCATGTCGT TAATTCTATG GTGTCTGATA TTGAAAGGGT  
1401 TGATTGAAC AGCATTGGCT CTCAAGCCCT TCTTACCATT TCAGGCAGCA  
10 1451 CAGACACTGA ATCTGGTATC TACACTGAAC ATGACTCTAT CACTGTGTT  
1501 CAAGAACTGA AGGATCTCCT TAAGAAGAAT GCCACTGTGG AGGCTTTAT  
1551 TGAATGGTTG GATACTGTGG TAGAACAGAG AGTTATTAAG ACCAGCAAAC  
1601 AAAATGGAAG GTCATTAAG AAGAGAGCTC AAGACTTTCT GTTAAAGTGG  
1651 AGTTTTTTG GTGCTCGAGT AATGCATAAT CTCACCTTGA ACAATGCATC  
20 1701 CAGTTTGTT TCTTTTCATT TGATTGAAT GCTTCTCGAT GAATACATT  
1751 TCCTGGCCAT GGAGACCCAG TTTAATAATG ACAAAAGAGCA GGAGTTACAG  
1801 AATTATTGG ACAAGTATAT GAAGAATTCA GATGCGAGTA AAGCTGCTT  
25 1851 CACTGCTTCT CCGAGTTCAT GCTTCTGGC CAACCGTAAT AAAGGGAGCA  
1901 TGGTTTCCAG CGACGCTGTG AAGAATGAAA GCCACGTGGA GACAACCTAT  
30 1951 CTCCCTCTGC CATCCAGTCA ACCTGGAGGC CTAGGCCCTG CTCTGCACCA  
2001 GTTCCCTGCT GGGAACACAG ACAACATGCC GCTCACAGGT CAAATGGAGC  
2051 TTTCACAGAT TGCTGGTCAT CTGATGACAC CACCCATTTC TCCAGCCATG  
35 2101 GCAAGCCGAG GAAGTGTCAAT TAACCAAGGA CCAATGGCAG GGAGGCCCC  
2151 AAGTGTGGGC CCAGTACTGT CAGCTCCATC ACACTGCTCC ACATAACCCAG  
40 2201 AGCCCATTAA TCCCACCTCTC CCTCAAGCCA ATCATGACTT TTATAGCACC  
2251 AGCTCTAACT ACCAGACTGT GTTTAGGGCA CAGCCCCACT CCACATCAGG  
2301 ACTCTATCCT CATCACACCG AGCATGGTCG ATGCATGGCT TGGACTGAAC  
45 2351 AGCAGCTTTC AAGAGACTTC TTCAGTGGCA GCTGTGCGGG GTCTCCATAT  
2401 AACTCCCGGC CACCGTCTAG CTATGGCCA TCCCTGCAAG CCCAGGATTC  
50 2451 ACACAATATG CAGTTTTAA ATACAGGAAG CTTCAATTTC TTGAGCAACA  
2501 CAGGAGCTGC CAGCTGCCAA GGAGCAACAC TGCCTCCTAA TTCACCAAAT  
2551 GGATACTATG GAAGCAACAT AAACTACCCA GAGTCTCACA GGCTCGGATC  
55 2601 AATGGTGAAT CAGCACGTTT CTGTCATCAG CAGCATTGCGT TCACTGCC

2651 CCTACAGTGA CATCCACGAT CCACTTAACA TTTTAGATGA CAGTGGTAGA  
 2701 AAACAGACCA GCTCGTTTA CACAGACACA TCATCTCCAG TTGCATGTCG  
 5 2751 AACTCCAGTC CTAGCTCCA GTTGCAAAC CCCAATTCCCT TCTTCCTCAT  
 2801 CCCAATGTAT GTATGGAAC TCCAACCAGT ATCCAGCTCA AGAAACCCCTG  
 10 2851 GACTCCCAG GAACAAGCAG TAGAGAAATG GTGTCCTCTT TACCACCTAT  
 2901 CAACACTGTG TTCATGGAA CAGCAGCTGG AGGCACCTAA ACCACCAATG  
 15 2951 TGGGAGGGGG TGCTAAAAT TTAAAAAAA TCTCTACTGT GCAAATATCA  
 3001 TTATTCACTC AGACTTCCAT AAGAGTAAAT AAAAATGAA TATGCAGTSEQ ID

NO:2

Human IC-RFX polypeptide sequence.

20 1 MAKVPELEDT FLQAQPAPQL SPGIQEDCCV QLLGKGLLVY PEETVYLAEE  
 51 51 GQPGGEQGGG EKGEDPELPG AVKSEMHLNN GNFSSEEEADA DNHDSKTCAA  
 101 101 DQYLSQKKTI TQIVKDKKKQ TQLTLQWLEE NYIVCEGVCL PRCILYAHYL  
 25 151 DFCRKEKLEP ACAATFGKTI RQKFPLLTTR RLGTRGHSKY HYYGIGIKES  
 201 201 SAYYHSVYSG KGLTRFSGSK LKNEGGFTRK YSLSSKTGTL LPEFPSAQHL  
 30 251 VYQGCISKDK VDTLIMMYKT HCQCILDNAI NGNFEEIQHF LLHFWQGMPD  
 301 301 HLLPLLENPV IIDIFCVCD S ILYKVLTDSL IPATMQEMPE SLLADIRNFA  
 35 351 KNWEQWVVSS LENLPEALTD KKIPIVRRFV SSLKRQTSFL HLAQIARPAL  
 401 401 FDQHVVNSMV SDIERVDLNS IGSQALLTIS GSTDTESGIY TEHDSITVFQ  
 451 451 ELKDLLKKNA TVEAFIEWLD TVVEQRVIKT SKQNGRSLKK RAQDFLLKWS  
 501 501 FFGARVMHNL TLNNASSFGS FHLIRMLDE YILLAMETQF NNDKEQELQN  
 551 551 LLDKYMKNSD ASKAAFTASP SSCFLANRNK GSMVSSDAVK NESHVETTYL  
 601 601 PLPSSQPGGL GPALHQFPAG NTDNMPLTGQ MELSQIAGHL MTPPISPAMA  
 651 651 SRGSVINQGP MAGRPPSVGP VLSAPSHCST YPEPIYPTLP QANHDFYSTS  
 701 701 SNYQTVFRAQ PHSTSGLYPH HTEHGRCMAW TEQQQLSRDFF SGSCAGSPYN  
 751 751 SRPPSSYGPS LQAQDSHNMQ FLNTGSFNFL SNTGAASCQG ATLPPNPN  
 801 801 YYGSNINYPE SHRLGSMVNQ HVSVISSLRS LPPYSDIHDP LNILDSSGRK  
 851 851 QTSSFYTDTS SPVACRTPVL ASSLQTPIPS SSSQCMYGTS NQYPAQETLD  
 901 901 SHGTSSREMV SSLPPINTVF MGTAAGGT

SEQ ID NO:3

RFX CONSENSUS SEQUENCE

XXPGXXXGGXXXXXRXXXXXXXEX  
5 XLNXGXXSEXXXXXXHSXXXXXXKXXXXXXXTXLQWLЕНYXXXEGVCLPRCXLYXHYLDF  
CKXXXXPXXAAXFGKIRQXFPLTTTRRLGTRGSKYHYYGIXXKESSXYXXYSXKGXXXSXXXXX  
XXYSXXSKGTLLPEFPXXQHXXXXXIXXXVXTLIMMYXTHCQXILDXXIXXNFEXQFLLHF  
LXXXXXXDIXXVCDSLYKXXXDVLI.PXXXQEXPXSLXXXIRXFAKNXXWXXSLXNLPEXL  
XXLKRQTSXXHLAQXXRLXXXXVXXMXXRVDLNSIXXQALXTXXSDXXXXXTXXXXXXQEXX  
10 XXLXXXXXXIEWLDTXXQVXXXXXXSLKXXAQXFLLWSFFGXRVXXLTLLXXAXSF  
15 GSFHLLIRXLXDXYX  
LXXXXXXQEXXNLXXXXMXXXXXGXXXXXXXSPXXXXXXPXXXXXX  
XXXXXXSXXXXXXSXXXXXXSXXXXXXSXXXXXXSXXXXXXSXXXXXXSXXXXXX  
XXXXXXSXXXXXXSNTGXXXXXXNGXXXXXXSXXXXXXSXXXXXX  
XXXXXXSXXXXXXSXXXXXXSXXXXXXSXXXXXXSXXXXXXSXXXXXX  
XXXXXXSXXXXXXFW

SEQ ID NO:4

DBD consensus

20 TLQWLXXNYXXEGVXLPRXXLYXHYLXXCXXXLEPXXAAXFGKIRXXFXXLXTRRLGTRGS  
KYHYYGIXX

SEQ ID NO:5

B domain consensus

25 VXXLXXXYYXXHCXXILDXXXNXXFXXXXXXXFW

SEQ ID NO:6

C domain consensus

30 LYXXXXXXLIPXXXXXXPXXLXXXIRXFAKXXXXWXXXL

SEQ ID NO:7

Dimerization domain

35 FXXXLXRXTSXHLAQXARXXLXXXXXXNMXSDXXRVDXNXXXQAXXXXXXXQXXXXXX  
XLSXXXXXXEWEWLDXVXXQXXXXXXYXXXXXXKAXXFLLKWSFXXXXVXXLTLLXXAXSF  
GSFHLIRXLXDE  
YXXXXXXSXXXXXXL